

Achievable Error Exponents in the Gaussian Channel with Rate-Limited Feedback

Reza Mirghaderi, Andrea Goldsmith, Tsachy Weissman

Abstract

We investigate the achievable error probability in communication over an AWGN discrete time memoryless channel with noiseless delay-less rate-limited feedback. For the case where the feedback rate R_{FB} is lower than the data rate R transmitted over the forward channel, we show that the decay of the probability of error is at most exponential in blocklength, and obtain an upper bound for increase in the error exponent due to feedback. Furthermore, we show that the use of feedback in this case results in an error exponent that is at least R_{FB} higher than the error exponent in the absence of feedback. For the case where the feedback rate exceeds the forward rate ($R_{FB} \geq R$), we propose a simple iterative scheme that achieves a probability of error that decays doubly exponentially with the codeword blocklength n . More generally, for some positive integer L , we show that a L^{th} order exponential error decay is achievable if $R_{FB} \geq (L - 1)R$. We prove that the above results hold whether the feedback constraint is expressed in terms of the average feedback rate or per channel use feedback rate. Our results show that the error exponent as a function of R_{FB} has a strong discontinuity at R , where it jumps from a finite value to infinity.

I. INTRODUCTION

While feedback cannot increase the capacity of a point-to-point memoryless channel, it can decrease the probability of error as well as the complexity of the encoder and decoder. For an AWGN channel without feedback, it is known [1] that the decay in the probability of error as a function of the blocklength n is at most exponential in the absence of feedback (i.e. the lowest achievable probability of error has the general form $P_e = \exp(-O(n))$).¹ However, when a noiseless delay-less infinite capacity feedback link is available, a simple sequential linear scheme (the Schalkwijk-Kailath scheme [2]) can achieve the capacity of this channel with a doubly exponential decay in the probability of error as a function of the blocklength (i.e. it has the general form $P_e = \exp(-\exp(\Omega(n)))$). This shows the significant role of feedback in reducing the probability of error.

The Schalkwijk-Kailath scheme requires a noiseless feedback link with infinite capacity. In fact, the Schalkwijk-Kailath scheme does not provide the best possible error decay rate given such an ideal feedback link. In particular, it is shown in [3] that in the presence of an ideal noise-free delay-less feedback link, the capacity of the AWGN channel can be achieved with a probability of error that decreases with an exponential order which is linearly increasing with blocklength (i.e. it has the general form $P_e = \exp(-\underbrace{\exp \circ \dots \circ \exp}_{\Omega(n) \text{ times}}(\Omega(n)))$).² However, once the feedback channel is corrupted with some

noise, the benefits of feedback in terms of the error probability decay rate can drop. In fact, when this corruption corresponds to an additive white Gaussian noise on the feedback channel, the Schalkwijk-Kailath communication scheme (or any other linear scheme) fails to achieve any nonzero rate with vanishing error probability [4]. Furthermore, in this case, the achievable error decay for any coding scheme can be no better than exponential in blocklength [5], similar to the case without feedback [1].

R. Mirghaderi, A. Goldsmith and T. Weissman are with the Department of Electrical Engineering, Stanford University, 350 Serra Mall, Stanford, CA, USA, email: rezam@stanford.edu, andrea@wsl.stanford.edu, tsachy@stanford.edu. This paper was presented in part at the 48th Annual Allerton Conference on Communications, Control and Computing, 2010. This work is supported by the NSF Center for the Science of Information under Award CCF-0939370.

¹Given a function $h(\cdot)$, $h(n) = O(n)$ is equivalent to $\lim_{n \rightarrow \infty} \frac{h(n)}{n} < \infty$, and $h(n) = \Omega(n)$ is equivalent to $\lim_{n \rightarrow \infty} \frac{h(n)}{n} > 0$.

²Operator \circ is used to denote function composition.

In this work, we consider a case where the feedback link is noiseless and delay-less but rate-limited. The advantages of rate-limited feedback in reducing the coding complexity are investigated in [6]. In this paper, we study the benefits of rate limited feedback in terms of decreasing the error probability. Assuming a positive and feasible (below capacity) rate R is to be transmitted on the forward channel, we characterize the achievable error decay rates in two cases: the case where the feedback rate, R_{FB} , is lower than R , and the case where $R_{FB} \geq R$. For the first scenario, we show that the best achievable error probability decreases exponentially in the code blocklength n (i.e. $P_e = \exp(-O(n))$) and provide an upper bound for the error exponent. For the second scenario, we propose an iterative coding scheme which achieves a doubly exponential error decay (i.e. $P_e = \exp(-\exp(\Omega(n)))$). Since a feedback rate equal to the data rate is sufficient for achieving a doubly exponential error decay, one might suspect that further increasing the feedback rate may not lead to a significant gain. We dispel this suspicion by generalizing our proposed iterative scheme to show that if $R_{FB} \geq (L-1)R$, an L^{th} order exponential decay is achievable. The latter result is consistent with [7], in which the achievable error probabilities are characterized in terms of the number of times the (infinite capacity) feedback link is used.

Interestingly, our results show that the error exponent as a function of the feedback rate has a strong discontinuity at the point $R_{FB} = R$; it is finite for $R_{FB} < R$ and infinite for $R_{FB} \geq R$ (due to the achievability of a doubly exponential error decay).

Although only $R_{FB} \geq R$ can lead to a super-exponential error decay, even for smaller feedback rates, we expect to have a strictly higher error decay rate as compared to the case with no feedback. In particular we show that for $R_{FB} < R$, the error exponent is at least R_{FB} higher than the error exponent in the absence of feedback.

The problem of communication over the AWGN channel with limited feedback has been previously considered assuming different types of corruption on the feedback channel. In particular, the corruption on the feedback channel has been modeled as additive Gaussian noise in [4] and [5] and as quantization noise in [8]. Another type of feedback corruption has been considered in [9] where only a subsequence of the channel outputs can be sent back noiselessly to the transmitter. A fundamental distinction between our model and the ones considered above is that in our model the receiver has “full control” over what is transmitted and received on the feedback link. This is due to the fact that under the rate-limited feedback scenario, the feedback link is assumed to be both noiseless and active in the sense that at each time, the feedback message is allowed to be an encoded function of all the information available at the receiver at that time. Communication with imperfect feedback has also been investigated in [10], [11] and [12] for variable-length coding strategies. Our model on the other hand captures a scenario where the blocklength and therefore the decoding delay is fixed.

The rest of this paper is organized as follows: In Section II we present the system model and the problem formulation. In Section III we consider the case where the feedback rate is higher than the forward rate. Specifically, using a simple iterative coding scheme we show the achievability of an L^{th} order exponential error decay when $R_{FB} \geq (L-1)R$. In Section IV we consider the case where $R_{FB} < R$ and show that in this case the decay in probability of error is at most exponential (finite first order error exponent). Although a feedback rate less than R cannot provide super-exponential error decay, we will show in Section V that it increases the error exponent by at least R_{FB} . Section VI shows that the necessary and sufficient conditions for super-exponential error decay remain the same even if we express the feedback limitation as a constraint on the per channel use feedback rate instead of the average feedback rate. Finally, Section VII concludes the paper.

Notation. Throughout this paper we represent the L_2 norm operator by $\|\cdot\|$ and the expectation operator by $E[\cdot]$. The notation “log” is used for the natural logarithm, and rates are expressed in nats. The complement of a set \mathcal{A} is denoted by \mathcal{A}^c . We denote the indicator function of the event \mathcal{A} by $1_{\mathcal{A}}$. Given a function $h(\cdot)$, $h(n) = o(1)$ is equivalent to $\lim_{n \rightarrow \infty} |h(n)| = 0$. Given a function $h(\cdot)$ and a positive integer k , the k^{th} iterate of the function, i.e. $\underbrace{h \circ \dots \circ h(\cdot)}_{k \text{ times}}$, is denoted by $h^k(\cdot)$.

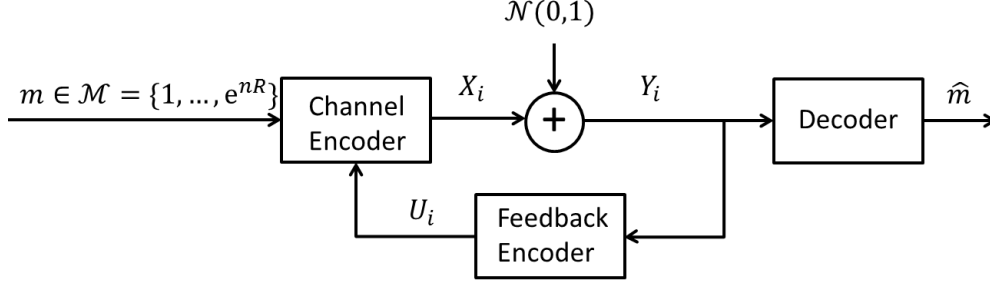


Fig. 1. AWGN channel with rate-limited feedback

II. SYSTEM MODEL

We consider communication over a block of length n through an AWGN channel with rate-limited noiseless feedback. The channel output Y_i at time i is given by

$$Y_i = X_i + N_i,$$

where $\{N_i\}_{i=1}^n$ is a white Gaussian noise process with $N_i \sim \mathcal{N}(0, 1)$ and X_i is the channel input at time i . The finite-alphabet feedback signal at time i is denoted by $U_i \in \mathcal{U}_i$ and is assumed to be decoded at the transmitter (of the forward channel) without any error or delay. We will denote the feedback sequence alphabet $\mathcal{U}_1 \times \dots \times \mathcal{U}_n$ by \mathcal{U} . The message m to be transmitted (on the forward link) is assumed to be drawn uniformly from the set $\mathcal{M} = \{1, \dots, |\mathcal{M}|\}$.

An encoding strategy is comprised of a sequence of functions $\{f_i^{(n)}\}_{i=1}^n$ where $f_i^{(n)} : \mathcal{M} \times \mathcal{U}_1 \times \dots \times \mathcal{U}_{i-1} \mapsto \mathbb{R}$ determines the input X_i as a function of the message and the feedback signals received before time i ,

$$X_i = f_i^{(n)}(m, U_1, \dots, U_{i-1}).$$

The feedback strategy consists of a sequence of functions $\{g_i^{(n)}\}_{i=1}^n$ where $g_i^{(n)} : \mathbb{R}^i \mapsto \mathcal{U}_i$ determines the feedback signal as a function of the channel outputs up to time i ,

$$U_i = g_i^{(n)}(Y_1, \dots, Y_i).$$

The decoding function $\phi : \mathbb{R}^n \mapsto \mathcal{M}$ gives the reconstruction of the message after receiving all the channel outputs

$$\hat{m} = \phi^{(n)}(Y^n).$$

The probability of error for message m is denoted by $P_e(m)$, where

$$P_e(m) = \Pr\{\hat{m} \neq m | m \text{ is transmitted}\}.$$

The average probability of error is defined as

$$P_e = \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} P_e(m).$$

Given the above setup, a communication scheme with forward rate R , feedback rate R_{FB} and power level P is comprised of a selection for the feedback sequence alphabet \mathcal{U} , the encoding strategy $\{f_i^{(n)}\}_{i=1}^n$, the feedback strategy $\{g_i^{(n)}\}_{i=1}^n$ and the decoding function $\phi^{(n)}(\cdot)$, such that

$$\begin{aligned} |\mathcal{M}| &\geq e^{nR}, \\ |\mathcal{U}| &\leq e^{nR_{FB}}, \end{aligned}$$

$$E\left[\sum_{i=1}^n \left(f_i^{(n)}(m, U^{i-1})\right)^2\right] \leq nP,$$

where the expectation is with respect to the messages and the noise. Over all such communication schemes, we represent the one with minimum average probability of error with the tuple (n, R, R_{FB}, P) and denote the corresponding minimum error probability by $P_e(n, R, R_{FB}, P)$. In the case where the feedback rate is zero, we simply drop the feedback rate term and use (n, R, P) and $P_e(n, R, P)$ to represent the optimal non-feedback code and the corresponding error probability, respectively. The capacity of the AWGN channel is denoted by C , where

$$C = \frac{1}{2} \log(1 + P).$$

For the communication system described above, the first order error exponent or simply the error exponent is defined as

$$E_1(R, R_{FB}, P) = \overline{\lim}_{n \rightarrow \infty} \frac{-\log P_e(n, R, R_{FB}, P)}{n}, \quad (1)$$

where a positive value of the error exponent implies that the error decay rate is at least exponential. We also define higher order error exponents. In particular, given $L \geq 2$, the L^{th} order error exponent is defined as

$$E_L(R, R_{FB}, P) = \overline{\lim}_{n \rightarrow \infty} \frac{\log^{L-1}(-\log P_e(n, R, R_{FB}, P))}{n}. \quad (2)$$

Given the above definitions, a communication system with strictly positive L^{th} order error exponent has an L^{th} order exponential error decay (i.e. $P_e(n, R, R_{FB}, P) = \exp(-\exp^{L-1}(\Omega(n)))$).

III. $R_{FB} \geq R$: SUPER-EXPONENTIAL ERROR DECAY

When the feedback rate is higher than the forward rate R , we can achieve a super-exponential (in blocklength) error decay. This result is presented in the following theorem.

Theorem 1 *For any $R > 0$ which satisfies $R \leq R_{FB}$ and $R < C$, a strictly positive second order error exponent is achievable:*

$$E_2(R, R_{FB}, P) > 0.$$

Proof: See Appendix. ■

The above result can be further generalized as follows.

Theorem 2 *Given an integer $L \geq 2$, for any $R > 0$ which satisfies $R \leq \frac{1}{L-1} R_{FB}$ and $R < C$, a strictly positive L^{th} order error exponent is achievable:*

$$E_L(R, R_{FB}, P) > 0.$$

Proof: See Appendix. ■

We use a class of simple iterative coding schemes to prove the above achievability results. In particular, to achieve a doubly exponential error decay we propose a multi-phase coding scheme as follows: in the first phase, called the initial transmission, the message is sent using a non-feedback code that occupies a big portion of the transmission block (n_1 out of n). In the second phase, called the intermediate decoding/feedback phase, the receiver decodes the message based on the received signals and feeds back the decoded message to the transmitter, using nR nats of the available feedback. Depending on the validity of the decoded message the transmitter decides to stay silent or perform boosted retransmission. In the case the message is decoded correctly, the transmitter stays silent during the rest of the transmission time. Otherwise, it sends a sign of failure in the next ($n_1 + 1^{st}$) transmission and uses the remaining portion of the transmission block ($n_2 = n - n_1 - 1$) to send the message with an exponentially (in block length)

high power. While retransmission with such a large power guarantees a doubly exponential error decay, it does not violate the power constraint since the probability of incorrect decoding in the second phase is exponentially (in block length) low.

To guarantee an L -fold exponential decay when the available feedback rate is $(L-1)R$, for some integer $L > 2$, the above scheme can be modified to include $L-1$ rounds of intermediate decoding/feedback and boosted retransmission, where retransmission at each round, if needed, is done with exponentially higher power than the previous retransmission.

Note that in comparison with the Schalkwijk-Kailath (SK) scheme presented in [2], the above iterative technique needs less feedback (LR nats instead of the infinite rate required by the SK scheme) and provides better error decay rate.

IV. $R_{FB} < R$: FIRST ORDER EXPONENTIAL ERROR DECAY

In the previous section we have shown that by utilizing a feedback link with a rate higher than the forward rate, we can reduce the error probability significantly as compared to the case with no feedback. The high reliability of the iterative scheme presented in the last section is due to the fact that the initial decoding error at the receiver (which is a rare event) is perfectly detectable at the transmitter. Therefore it can be corrected by retransmitting the message with high power without violating the average power constraint. The perfect error detection at the transmitter is obtained from the feedback of the initial decoded message at the receiver. However, when the feedback rate is lower than the forward rate, the receiver has to use a source code to compress its decoded message before feeding it back. The transmitter must then reconstruct the uncompressed decoded message to detect any error. Since this reconstruction involves some first order exponential (in blocklength) error decay (corresponding to the source coding error exponent), the error detection is erroneous with the same decay rate. Therefore, the mis-detection of the receiver error due to the compression on the feedback link dominates the error probability.

While the above intuitive explanation justifies the failure of the block retransmission schemes in achieving a super-exponential error decay, one might still hope that such a decay rate can be achieved using other schemes. For example one alternative is to look at the problem from a stochastic control point of view and use a rate-limited variant of the recursive feedback schemes presented in [13] and [14]. In this section, we show that no matter what communication scheme is used, one cannot achieve infinite first order error exponent.

Theorem 3 *Given $R > R_{FB}$, the first order error exponent is upper bounded by*

$$E_1(R, R_{FB}, P) \leq E_{up}(R_{FB}),$$

where $E_{up}(R_{FB}) = 4P + \tau_0/2 + R_{FB}$ and τ_0 is the solution to $\frac{1}{2}(\tau_0 - 1 - \log(\tau_0)) = R_{FB}$.

Proof: See Appendix. ■

The proof, which is rather lengthy, can be explained using the following observation. It is shown in [15] that given a peak power constraint, the best achievable error decay is exponential. Therefore, in order to achieve a super-exponential error decay, the transmitter should be able to boost the power under certain circumstances. However, given the expected power constraint, the power can be boosted only under rare occasions where the receiver would decode wrongly otherwise. Therefore, there should be enough feedback bits to communicate the occurrence of those rare occasions to the sender. It turns out that this requirement is met only if the number of possible feedback messages ($e^{nR_{FB}}$) is at least as large as the number of forward messages (e^{nR}).

Note that the error exponent upper bound provided in the above theorem stays bounded as R_{FB} approaches R from below. On the other hand, we showed in the previous section that for any feedback rate higher than R , the error exponent is infinite (doubly exponential decay). These two facts lead to an interesting conclusion: the error exponent as a function of the feedback rate has a sharp discontinuity at the point $R_{FB} = R$.

The above theorem provides an upper bound on the first order error exponent for feedback rates below R . We conjecture that a similar result may be obtained on the boundedness of the L^{th} order error exponent for feedback rates below LR .

V. $R_{FB} < R$: LOWER BOUND ON ERROR EXPONENT

We have shown in the previous section that the probability of error when $R_{FB} < R$ cannot decay faster than exponential as a function of the blocklength n . Although the feedback in this case does not provide an infinite error exponent, we still expect that the error exponent should be improved in the presence of feedback as compared to the non-feedback scenario. In this section we will show that the error exponent with feedback is at least R_{FB} above the non-feedback error exponent. The main result of this section is the following theorem.

Theorem 4 *For all rates $R < C$, such that $R > R_{FB}$, the error exponent is lower bounded as follows*

$$E_1(R, R_{FB}, P) \geq E_{NoFB}(R) + R_{FB}, \quad (3)$$

where $E_{NoFB}(R)$ is the error exponent for the AWGN channel in the absence of feedback.

Proof: See Appendix. ■

The achievability scheme for the above result is constructed using the multi-phase scheme proposed in the proof of Theorem 1, in conjunction with a compression technique to reduce the rate of feedback in the intermediate decoding/feedback phase from R to R_{FB} . Using such a scheme, the error probability is dominated by the probability of error mis-detection. This error term is the product of the probability of error in the initial transmission phase ($\exp(-nE_{NoFB}(R))$) and the probability ($\exp(-nR_{FB})$) that the compression loss hides this event from the transmitter.

VI. PER CHANNEL USE FEEDBACK CONSTRAINT

In the previous sections we focused on a scenario where the *average* rate over the whole transmission block was constrained to be lower than R_{FB} . Under that constraint, the receiver can use the available feedback (nR_{FB} nats) any time during the transmission. In particular, using the coding scheme proposed in Section III, the receiver collects all the feedback bits and uses them in one feedback transmission at the end of the first phase. In this section we consider a *per channel use* feedback rate constraint. Under this constraint, the receiver cannot feed back more than R_{FB} nats after each channel use. This translates to the following constraint on the size of the feedback signal alphabet at each time $i \in \{1, \dots, n\}$:

$$|\mathcal{U}_i| \leq e^{R_{FB}}. \quad (4)$$

Given that the above constraint is more restrictive than the average feedback rate constraint considered previously, we can conclude that the upper bound on the error exponent obtained in Section IV holds in the above scenario as well. Interestingly, we show that similar achievability results as those stated in Section III for the average feedback rate constraint are also true for the per channel use feedback scenario.

Theorem 5 *Given the per channel use feedback constraint, if $R_{FB} \geq (L - 1)R$ and $R < C$, a strictly positive L^{th} order error exponent is achievable:*

$$E_L(R, R_{FB}, P) > 0.$$

Proof: See Appendix. ■

The above result is proved using a combination of the scheme presented in Section III and a block Markov coding scheme which is described in the Appendix. Figure 2 illustrates an example of this iterative coding scheme for the case where $L = 2$.

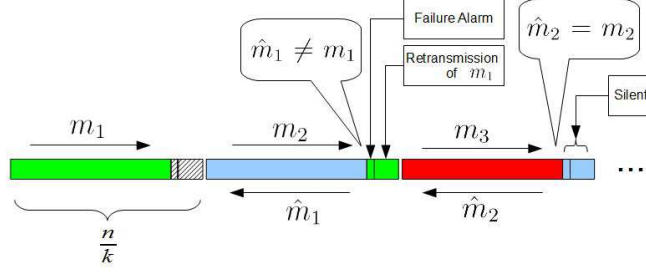


Fig. 2. Iterative feedback scheme for per channel use feedback constraint: An example

VII. SUMMARY AND DISCUSSION

We considered the impact of rate-limited noiseless feedback on the error probability in AWGN channels. We first showed that if the feedback rate R_{FB} that exceeds the rate R of the data transmitted on the forward channel, one can achieve a super-exponential decay in probability of error as a function of the code blocklength. Our achievability result is based on a multi-phase scheme in which an initial transmission of the message, if decoded incorrectly, is followed by the retransmission of the message with boosted power. A key requirement in this scheme is for the transmitter to perfectly detect the error in the initial transmission every time it happens. The minimum feedback rate required to perfectly communicate the initial decoded message is R and therefore our scheme fails to achieve a super-exponential error decay for $R_{FB} < R$. We showed that this is true for any scheme. That is, $R_{FB} \geq R$ is also a necessary condition for achieving a super-exponential error decay. While we provided an upper bound for the error exponent when $R_{FB} < R$, we also showed that even in this case, the use of feedback increases the error exponent by at least R_{FB} . For the case in which $R_{FB} \geq (L-1)R$, for some positive integer L , we generalized our multi-phase iterative scheme to prove the achievability of an L -fold exponential (in blocklength) error decay. The above results are illustrated in Figure 3. It can be seen that the error exponent as a function of the feedback rate has a sharp discontinuity at $R_{FB} = R$.

We showed that the above necessary and sufficient condition for achieving a super-exponential error decay holds whether the feedback limitation is expressed as a constraint on the *average* feedback rate or on the *per channel use* feedback rate. Note that our results address the asymptotic behavior of the probability of error in terms of the blocklength n and therefore may provide limited insight for codes with small blocklength. In particular, for small values of n , one might expect the per channel feedback rate constraint to lead to a higher error probability than a scenario with average feedback rate constraint. On the other hand, the former is a more practical scenario as it implicitly captures the delay associated with sending data on the feedback link.

In this paper we showed the advantages of feedback in terms of improving the decay rate of the error probability. A subject for future research is to explore the other advantages of interactive communication in terms of reducing the coding complexity and energy consumption. One interesting problem to be addressed is how to use rate-limited feedback to construct SK like schemes which do not need complex block encoding decoding.

VIII. APPENDIX

Proof of Theorem 1: Fix $\delta > 0$ such that $R < C(1 - \delta)$. Define $n_2 = \epsilon n$ and $n_1 = n - n_2 - 1$, where $\epsilon > 0$ is chosen such that

$$\frac{n}{n_1} < 1 + \delta \quad (5)$$

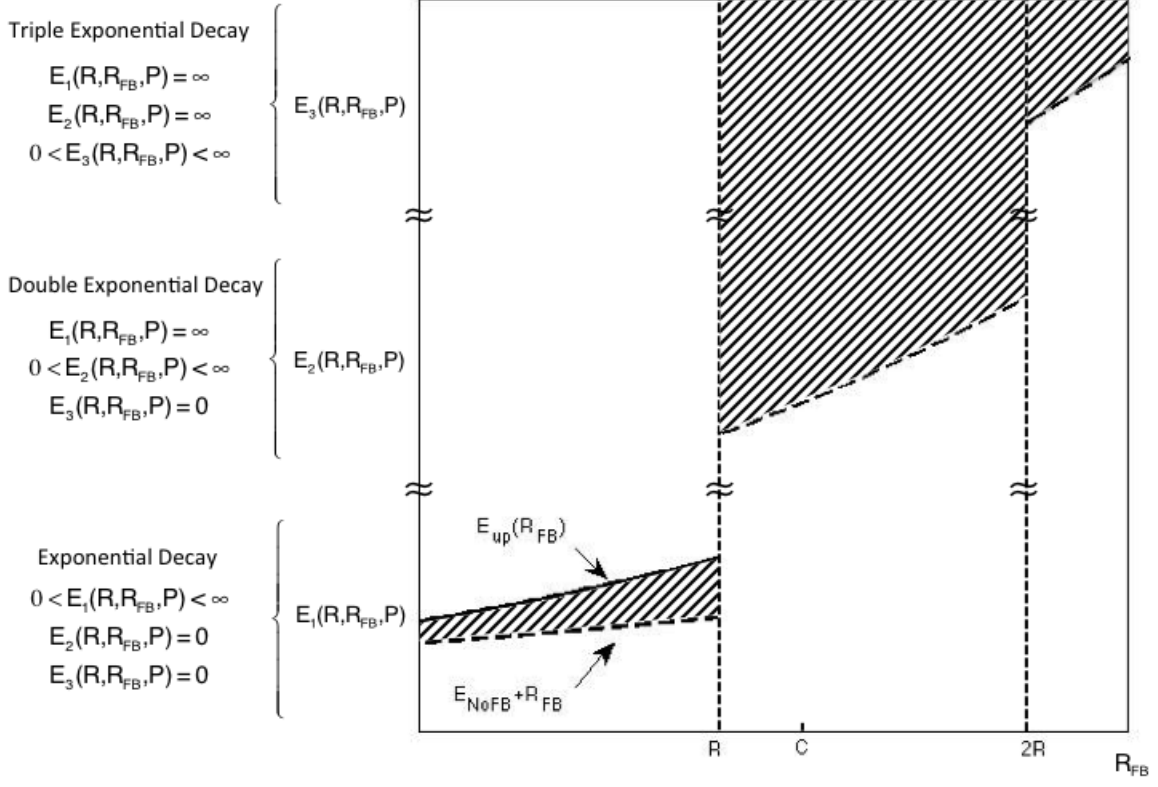


Fig. 3. An illustration of the bounds on error exponents in terms of the feedback rate R_{FB}

holds for large enough n . Choose the feedback signal domains as follows

$$\begin{aligned}\mathcal{U}_i &= \{1\}, \text{ for } i \neq n_1 \\ \mathcal{U}_{n_1} &= \{1, \dots, e^{nR}\}\end{aligned}$$

We construct two non-feedback codes $\mathcal{C}_1 = (n_1, \frac{nR}{n_1}, P)$ and $\mathcal{C}_2 = (n_2, \frac{nR}{n_2}, P/\gamma)$, where

$$\gamma = P_e \left(n_1, \frac{nR}{n_1}, P \right). \quad (6)$$

For $m \in \{1, \dots, e^{nR}\}$, pick the corresponding codeword $X^{n_1}(m)$ from \mathcal{C}_1 and send it in the first n_1 channel uses. Based on the received signals Y^{n_1} and using the optimal non-feedback decoding function for code \mathcal{C}_1 , the transmitter decodes the message and sends back its decision \hat{m}_1 to the transmitter

$$U_{n_1} = \hat{m}_1.$$

If $\hat{m}_1 = m$, then

$$X_i = 0, i = n_1 + 1, \dots, n,$$

otherwise, the next input will be

$$X_{n_1+1} = \sqrt{P/\gamma}$$

and then the codeword corresponding to m is picked from the codebook \mathcal{C}_2 and is transmitted in the remaining n_2 transmissions. On the other side, the receiver compares Y_{n_1+1} with the threshold $\Gamma = \frac{\sqrt{P/\gamma}}{2}$.

If $Y_{n_1+1} < \Gamma$, then the remaining received signals are ignored and the decoded message in the first try is announced as the final decision

$$\hat{m} = \hat{m}_1.$$

If $Y_{n_1+1} \geq \Gamma$, the receiver decodes the message based on the last n_2 received signals and using the optimal non-feedback decoding function for code \mathcal{C}_2 . The resulting message \hat{m}_2 is then announced as the final decision

$$\hat{m} = \hat{m}_2.$$

Using the above scheme, the average power used in the forward link will be

$$\frac{1}{n} (n_1 P + \gamma(n_2)(P/\gamma)) < P.$$

Therefore our scheme satisfies the power constraint. Also the average feedback rate is R which meets the constraint on the feedback link. There are three cases in which an error can happen. The first case is when the first decoding is correct but the receiver receives a failure signal from the transmitter due to the noise on the $n_1 + 1^{st}$ transmission. The probability of this event is upper bounded by

$$P_e\{\text{false negative}\} \leq Q(\Gamma), \quad (7)$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution. The second case is when the first decoding is wrong but the failure signal is not decoded correctly at the receiver. The probability of this event is upper bounded by

$$P_e\{\text{false positive}\} \leq Q(\Gamma). \quad (8)$$

The third case is when the first decoding fails and the failure signal is decoded correctly, but the second decoding also fails. The probability of this event satisfies

$$P_e\{\text{wrong decoding}\} \leq P_e(n_2, \frac{nR}{n_2}, P/\gamma) \quad (9)$$

$$= P_e(n_2, \frac{R}{\epsilon}, P/\gamma). \quad (10)$$

Using the exponential upper bound for the Q -function, we have

$$P_e\{\text{false negative}\} + P_e\{\text{false positive}\} \leq \alpha \exp(-\frac{P}{8\gamma}), \quad (11)$$

where $\alpha > 0$ is some constant. By positivity of the error exponent for rates less than the capacity [1] and since $\frac{nR}{n_1} \leq C(1 - \delta^2)$, we know that for any $\delta > 0$, there exists a fixed $\zeta > 0$ such that

$$\gamma = P_e\left(n_1, \frac{nR}{n_1}, P\right) \leq e^{-n\zeta}. \quad (12)$$

for large enough values of n . Combining (11) and (12), we obtain

$$P_e\{\text{false negative}\} + P_e\{\text{false positive}\} \leq \exp(-e^{n(\zeta+o(1))}), \quad (13)$$

which shows the probability of the first two types of errors decays doubly exponentially in the blocklength. It remains to show that the third type of error is also upper bounded by a doubly exponential term. To show that, note that on the right hand side of (10), the rate is at most $1/\epsilon$ times the capacity achieved by SNR P . However, the SNR P/γ is exponentially (in n) higher than P

$$P/\gamma \geq P e^{n\zeta},$$

for large values of n and therefore

$$P_e\{\text{wrong decoding}\} \leq P_e(\epsilon n, \frac{R}{\epsilon}, P e^{n\zeta}). \quad (14)$$

Given (13) and the above inequality, the proof will be complete if we show that $P_e(\epsilon n, \frac{R}{\epsilon}, P e^{n\zeta})$ decays doubly exponentially as a function of n . To show this, we can use the fact that for communication rates (in nats/channel use) less than

$$\frac{1}{2} \ln \frac{2 + \sqrt{P^2 + 4}}{4},$$

the following upper bound on error probability holds in the absence of feedback [1]:

$$P_e(n, R, P) \leq e^{-n(E(R, P) - \epsilon')},$$

for any $\epsilon' > 0$ and for large enough values of n , where

$$E(R, P) = \frac{P}{4} (1 - \sqrt{(1 - e^{-2R})}). \quad (15)$$

Take n sufficiently large such that

$$\frac{R}{\epsilon} < \frac{1}{2} \ln \frac{2 + \sqrt{P^2 e^{2n\zeta} + 4}}{4},$$

i.e.

$$n \geq \frac{1}{\zeta} \ln \frac{(4e^{2\frac{R}{\epsilon}} - 2)^2 - 4}{P^2}.$$

Then using (15) leads to

$$\begin{aligned} P_e(\epsilon n, R/\epsilon, P e^{n\zeta}) &\leq e^{-n\epsilon(\frac{P e^{n\zeta}}{4}(1 - \sqrt{1 - e^{-\frac{2R}{\epsilon}}}) + \epsilon')} \\ &= \exp(-\exp(n(\zeta + o(1)))) \end{aligned}$$

■

Proof of Theorem 2: Let's partition the whole transmission block into $L + 1$ sub-blocks, the first of which has length $(1 - \epsilon)n$. We choose the remaining sub-blocks to have equal lengths. In the first sub-block, the transmitter sends the message using the non-feedback Gaussian codebook \mathcal{C}_1 with rate R and power P . After transmission in the i^{th} sub-block, the receiver feeds back the message it has decoded within that sub-block. If the decoded message matches the transmitted one, the transmitter stays silent for the rest of the time. Otherwise, it sends a failure alarm and retransmits the message in the $i + 1^{st}$ sub-block using a non-feedback Gaussian codebook \mathcal{C}_i with rate R . The power of the alarm signal and the power P_i of codebook \mathcal{C}_i are chosen to be inversely proportional to the probability of decoding error in the first i sub-blocks. That is,

$$P_{i+1} = P/\gamma_i,$$

where γ_i is the total probability of error in the first i sub-blocks. The L -fold exponential error decay can be shown inductively. Given that the probability γ_i is $(i - 1)$ -fold exponential in terms of the blocklength (the case of $i = 2$ was shown in the previous Theorem), the power at the i^{th} sub-block (if transmission is needed) is $(i - 1)$ -fold exponential in blocklength. This in turn leads to an i -fold exponential error decay at the end of the i^{th} sub-block. Note that both the transmission power and the feedback rate in the above scheme satisfy the problem constraints. ■

Proof of Theorem 3: Let us first introduce some key definitions which will be used in our proof. We define the decoding region for message m as

$$D(m) = \{Y^n : \phi^{(n)}(Y^n) = m\}$$

Also for each feedback signal sequence $u^n = (u_1, \dots, u_n) \in \mathcal{U}$, let's define the feedback decision region

$$B(u^n) = \{Y^n : g_i^{(n)}(Y^i) = u_i, i = 1, \dots, n\}.$$

A key quantity in our proof is the joint distribution of the feedback signal sequence and the output sequence given the transmitted message $P_{Y^n, U^n|M}(\cdot, \cdot|\cdot)$. For simplicity, we drop the subscript and use $P(y^n, u^n|m)$ to denote the density of the output sequence y^n and the feedback sequence $u^n = (u_1, \dots, u_n)$ conditional on the transmission of the message m . Defining $u_0 = 0$, we can write

$$P(y^n, u^n|m) = \prod_{i=1}^n P(y_i|m, u^{i-1}, y^{i-1}) P(u_i|m, u^{i-1}, y^i) \quad (16)$$

$$= \prod_{i=1}^n P(y_i|m, u^{i-1}, f_i^{(n)}(m, u^{i-1}), y^{i-1}) P(u_i|m, u^{i-1}, y^i, g_i^{(n)}(y^i)) \quad (17)$$

$$= \prod_{i=1}^n P(y_i|f_i^{(n)}(m, u^{i-1})) \mathbf{1}_{\{u_i = g_i^{(n)}(y^i)\}} \quad (18)$$

$$= \mathbf{1}_{\{y^n \in B(u^n)\}} \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{(y_i - f_i^{(n)}(m, u^{i-1}))^2}{2}\right) \quad (19)$$

$$= \mathbf{1}_{\{y^n \in B(u^n)\}} (2\pi)^{-n/2} \exp\left(-\frac{\|y^n - f^{(n)}(m, u^n)\|^2}{2}\right), \quad (20)$$

where $f^{(n)}(m, u^n) = (f_1^{(n)}(m, u_0), \dots, f_n^{(n)}(m, u^{n-1}))$. In this derivation, (16) is a consequence of the probability chain rule. Equation (17) is derived using the fact that for any two random variables (W, S) and any deterministic mapping $T(\cdot)$, $W \leftrightarrow S \leftrightarrow T(S)$ is a Markov chain. Finally, (18) is a direct result of the Markov chain relationship $(M, U^{i-1}, Y^{i-1}) \leftrightarrow X_i \leftrightarrow Y_i$ and also the equation $U_i = g_i^{(n)}(Y^i)$. Another quantity of interest will be the probability of using a feedback signal sequence $u^n \in \mathcal{U}$ conditional on the transmission of a message $m \in \mathcal{M}$,

$$P(u^n|m) = \int P(y^n, u^n|m) dy^n. \quad (21)$$

With the above definitions we can now proceed with the proof. Suppose the theorem does not hold. That is, let's assume there exists $\gamma > 0$ such that the following inequality can hold for arbitrarily large n :

$$P_e(n, R, R_{FB}, P) < e^{-n(E_{up}(R_{FB}) + \gamma)}. \quad (22)$$

Given such n 's, the above inequality implies that for at least half of the messages $m \in \mathcal{M}$, we have

$$P_e(m) < 2e^{-n(E_{up}(R_{FB}) + \gamma)} = e^{-n(E_{up}(R_{FB}) + \gamma + o(1))}. \quad (23)$$

Removing the messages which do not satisfy the above, we obtain a codebook with the rate of at least $\frac{1}{n} \log(\frac{e^{nR}}{2})$ which, for arbitrarily large n , is arbitrarily close to R . Therefore, (22) implies the existence of a code with rate R for which the *per message error probability* can be less than its right hand side for arbitrarily large n and for some $\gamma > 0$. Let us define $s(n) = n(E_{up}(R_{FB}) + \gamma)$. To prove the theorem, we will show that there exists n_0 such that for any $n > n_0$, the inequality

$$P_e(m) < e^{-s(n)} \quad (24)$$

cannot hold for all messages $m \in \mathcal{M}$. Let us fix n_0 , to be determined later, and assume that for some $n > n_0$, there exists a communication scheme for which (24) holds for all m . Given such a communication

scheme, for each m , we construct an initial bin $F_0(m)$ including a subset of feedback signal sequences as follows

$$F_0(m) = \{u^n : P(u^n|m) > \delta e^{-nR_{FB}}\},$$

where $\delta > 0$ is a fixed constant, to be determined later. Defining $\Pr\{F_0(m)|m\}$ as $\sum_{u^n \in F_0(m)} P(u^n|m)$, we can write

$$\begin{aligned} \Pr\{F_0(m)|m\} &= 1 - \sum_{u^n \notin F_0(m)} P(u^n|m) \\ &\geq 1 - \delta |\mathcal{U}| e^{-nR_{FB}} \\ &\geq 1 - \delta \end{aligned} \quad (25)$$

In the following algorithm we update the content of each bin sequentially.

- 1) Start with $i = 0$.
- 2) Pick two distinct messages $m, m' \in \mathcal{M}$, such that there exists a feedback sequence u^n where both $F_i(m)$ and $F_i(m')$ include u^n .
- 3) Assuming $\|f^{(n)}(m, u^n)\|^2 > \|f^{(n)}(m', u^n)\|^2$ (without loss of generality), remove u^n from $F_i(m)$.
- 4) Increase i by 1 and set $F_i(k) = F_{i-1}(k)$, for all $k \in \mathcal{M}$.
- 5) Set $J = \{k \in \mathcal{M} : F_i(k) \neq \emptyset\}$. If $|J| > e^{nR_{FB}}$, go to step 2, otherwise stop.

Note that step 2 is feasible since whenever this step is executed the number of non-empty bins are greater than the cardinality of $|\mathcal{U}|$ which is $e^{nR_{FB}}$. Therefore, there should exist at least one feedback sequence which appears in two bins. Also note that for any $k \in \mathcal{M}$ and any integer i

$$F_i(k) \subseteq F_{i-1}(k) \dots \subseteq F_0(k). \quad (26)$$

Assume m, m' are the messages picked in step 2 and u^n is the sequence removed from the bin $F_i(m)$ in step 3 and at iteration i of the above algorithm. Given such a 3-tuple (u^n, m, m') , a major part of the rest of the proof is devoted to obtaining a lower bound for $\|f^{(n)}(m, u^n)\|^2$. First for any y^n , let's use the triangle inequality to write

$$\begin{aligned} \|y^n - f^{(n)}(m, u^n)\|^2 &\leq (\|y^n - f^{(n)}(m', u^n)\| + \|f^{(n)}(m, u^n) - f^{(n)}(m', u^n)\|)^2 \\ &= \|y^n - f^{(n)}(m', u^n)\|^2 + \|f^{(n)}(m, u^n) - f^{(n)}(m', u^n)\|^2 \\ &\quad + 2\|y^n - f^{(n)}(m', u^n)\| \cdot \|f^{(n)}(m, u^n) - f^{(n)}(m', u^n)\| \\ &\leq 2(\|y^n - f^{(n)}(m', u^n)\|^2 + \|f^{(n)}(m, u^n) - f^{(n)}(m', u^n)\|^2). \end{aligned} \quad (27)$$

Similarly, we have

$$\|f^{(n)}(m, u^n) - f^{(n)}(m', u^n)\|^2 \leq 2(\|f^{(n)}(m, u^n)\|^2 + \|f^{(n)}(m', u^n)\|^2).$$

Combining (27), (28) and the assumption in step 3 of our algorithm that $\|f^{(n)}(m, u^n)\|^2 \geq \|f^{(n)}(m', u^n)\|^2$, we have

$$\|y^n - f^{(n)}(m, u^n)\|^2 \leq 2(\|y^n - f^{(n)}(m', u^n)\|^2 + 4\|f^{(n)}(m, u^n)\|^2).$$

Using this inequality and the derivation in (20), we have

$$P(y^n, u^n|m) > \mathbf{1}_{\{y^n \in B(u^n)\}} \exp(-4\|f^{(n)}(m, u^n)\|^2) (2\pi)^{-\frac{n}{2}} \exp(-\|y^n - f^{(n)}(m', u^n)\|^2). \quad (28)$$

Denoting the complement of a set \mathcal{A} by \mathcal{A}^c , we can write

$$P_e(m) = \int_{D(m)^c} \left(\sum_{u^n \in \mathcal{U}} P(y^n, u^n | m) \right) dy^n \quad (29)$$

$$\geq \int_{D(m)^c \cap B(u^n)} P(y^n, u^n | m) dy^n \quad (30)$$

$$\geq \int_{D(m') \cap B(u^n)} P(y^n, u^n | m) dy^n \quad (31)$$

$$\geq \exp(-4\|f^{(n)}(m, u^n)\|^2) \int_{D(m') \cap B(u^n)} (2\pi)^{-\frac{n}{2}} \exp(-\|y^n - f^{(n)}(m', u^n)\|^2) dy^n, \quad (32)$$

where (31) is due to the fact that $D(m)$ and $D(m')$ are disjoint sets and the last inequality is a consequence of (28). Using the assumption (24) and rearranging the above inequality, we can write

$$\|f^{(n)}(m, u^n)\|^2 \geq \frac{1}{4} \left(s(n) + \log \int_{D(m') \cap B(u^n)} (2\pi)^{-\frac{n}{2}} \exp(-\|y^n - f^{(n)}(m', u^n)\|^2) dy^n \right). \quad (33)$$

To complete our lower bound for $\|f^{(n)}(m, u^n)\|^2$, in the following, we find a lower bound for the integral in (33). First note that since $u^n \in F_i(m)$, we can write

$$\begin{aligned} & \int_{D(m') \cap B(u^n)} P(y^n, u^n | m') dy^n \\ &= P(u^n | m') - \int_{D(m')^c \cap B(u^n)} P(y^n, u^n | m') dy^n \\ &\geq P(u^n | m') - P_e(m') \\ &\geq \delta e^{-nR_{FB}} - e^{-s(n)} \end{aligned} \quad (34)$$

$$\begin{aligned} &\geq \delta e^{-nR_{FB}} \left(1 - \frac{1}{\delta} e^{-(s(n) - nR_{FB})}\right) \\ &\geq \frac{\delta}{2} e^{-nR_{FB}}, \end{aligned} \quad (35)$$

where (34) follows from the assumption that (24) holds for all the messages and the fact that u^n picked in step 3 and at the i^{th} iteration of the algorithm is in bin $F_i(m')$ and therefore is a member of $F_0(m')$. Also inequality (35) is secured by the appropriate choice of n_0 . Now let's define the sphere $Sp(f^{(n)}(m', u^n))$ as

$$Sp(m', u^n) = \{y^n : \|y^n - f^{(n)}(m', u^n)\|^2 \leq n\tau\}, \quad (36)$$

where τ will be determined later. Partitioning the set $D(m') \cap B(u^n)$ into $D(m') \cap B(u^n) \cap Sp(m', u^n)$ and $D(m') \cap B(u^n) \cap Sp(m', u^n)^c$ and using (35), we can write

$$\int_{D(m') \cap B(u^n) \cap Sp(m', u^n)} P(y^n, u^n | m') dy^n \geq \frac{\delta}{2} e^{-nR_{FB}} - \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)^c} P(y^n, u^n | m') dy^n. \quad (37)$$

The second term in the right hand side of (37) can be bounded as follows

$$\begin{aligned} & \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)^c} P(y^n, u^n | m') dy^n \\ &\leq \int_{Sp(m', u^n)^c} P(y^n, u^n | m') dy^n \\ &\leq \Pr \left\{ \sum_{i=1}^n (y_i - f_i^{(n)}(m', u^{i-1}))^2 \geq n\tau \right\} \\ &\leq \exp(-nE_c(\tau)), \end{aligned} \quad (38)$$

where we have used the Chernoff bound in the last step. In that inequality $E_c(\tau)$ is defined as

$$E_c(\tau) = \max_{s \geq 0} s\tau - \mu(s), \quad (39)$$

where $\mu(s)$ is the semi-invariant moment-generating function of the Chi-square distribution corresponding to $\kappa = (y_i - f_i^{(n)}(m', u^{i-1}))^2$:

$$\mu(s) = \log E_\kappa[e^{s\kappa}] = \frac{1}{2} \log\left(\frac{1}{1-2s}\right). \quad (40)$$

Replacing $\mu(s)$ in (39) and optimizing that equation we obtain

$$E_c(\tau) = \frac{1}{2}(\tau - 1 - \log(\tau)) \quad (41)$$

which is positive and increasing for all $\tau > 1$ and tends to infinity as $\tau \rightarrow \infty$. Choose τ such that

$$E_c(\tau) > R_{FB} + \epsilon, \quad (42)$$

for some $\epsilon > 0$, to be determined later. Using (37) and (38) we can write

$$\begin{aligned} & \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)} P(y^n, u^n | m') dy^n \\ & \geq \frac{\delta}{2} e^{-nR_{FB}} - e^{-n(R_{FB} + \epsilon)} \end{aligned} \quad (43)$$

$$\begin{aligned} & \geq \frac{\delta}{2} e^{-nR_{FB}} \left(1 - \frac{2}{\delta} e^{-n\epsilon}\right) \\ & \geq \frac{\delta}{4} e^{-nR_{FB}}, \end{aligned} \quad (44)$$

where we guarantee the validity of the last step by the appropriate choice of n_0 . Now let's derive the lower bound for the integral in (33) as follows

$$\int_{D(m') \cap B(u^n)} (2\pi)^{-n/2} \exp(-\|y^n - f^{(n)}(m', u^n)\|^2) dy^n \quad (45)$$

$$\geq \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)} (2\pi)^{-n/2} \exp(-\|y^n - f^{(n)}(m', u^n)\|^2) dy^n \quad (46)$$

$$\geq e^{-n\tau/2} \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)} (2\pi)^{-n/2} \exp\left(-\frac{\|y^n - f^{(n)}(m', u^n)\|^2}{2}\right) dy^n \quad (47)$$

$$= e^{-n\tau/2} \int_{D(m') \cap B(u^n) \cap Sp(m', u^n)} P(y^n, u^n | m') dy^n \quad (48)$$

$$\geq \frac{\delta}{4} e^{-n(\tau/2 + R_{FB})}. \quad (49)$$

The inequality (49) along with (33) lead to

$$\frac{\|f^{(n)}(m, u^n)\|^2}{n} \geq \frac{1}{4} \left(\frac{s(n)}{n} - \frac{\log(\frac{4}{\delta})}{n} - \frac{\tau}{2} - R_{FB} \right). \quad (50)$$

Substituting $s(n) = n(E_{up}(R_{FB}) + \gamma)$ in the above inequality, we obtain

$$\frac{\|f^{(n)}(m, u^n)\|^2}{n} \geq P + \frac{1}{4} \left(\gamma - \frac{\tau - \tau_0}{2} - \frac{\log(\frac{4}{\delta})}{n} \right). \quad (51)$$

By choosing ϵ in (42) small enough such that $\frac{\tau - \tau_0}{2} + \frac{\log(\frac{4}{\delta})}{n} < \gamma/2$, we conclude that for any feedback sequence u^n which is dropped in any iteration of our algorithm:

$$\|f^{(n)}(m, u^n)\|^2 > n(P + \frac{\gamma}{8}). \quad (52)$$

The above inequality is sufficient for us to prove the theorem. Noting that the cardinality of the set J at the end of our algorithm is $e^{nR_{FB}}$, we can write

$$E\left[\sum_{i=1}^n \left(f_i^{(n)}(m, U^{i-1})\right)^2\right] \quad (53)$$

$$= \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|} \sum_{u^n \in \mathcal{U}} P(u^n|m) \|f^{(n)}(m, u^n)\|^2 \quad (54)$$

$$\geq \sum_{m \in \mathcal{M} \setminus J} \frac{1}{|\mathcal{M}|} \sum_{u^n \in F_0(m)} P(u^n|m) \|f^{(n)}(m, u^n)\|^2 \quad (55)$$

$$\geq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M} \setminus J} \sum_{u^n \in F_0(m)} P(u^n|m) n(P + \frac{\gamma}{8}) \quad (56)$$

$$= \frac{n(P + \frac{\gamma}{8})}{|\mathcal{M}|} \sum_{m \in \mathcal{M} \setminus J} \Pr\{F_0(m)|m\}. \quad (57)$$

$$\geq \frac{n(P + \frac{\gamma}{8})}{|\mathcal{M}|} \sum_{m \in \mathcal{M} \setminus J} (1 - \delta) \quad (58)$$

$$\geq n(P + \frac{\gamma}{16})(1 - e^{-n(R-R_{FB})}) \quad (59)$$

$$> nP. \quad (60)$$

In the above derivation, (56) is obtained using (52) and the fact that for all $m \in \mathcal{M} \setminus J$, all the u^n 's in $F_0(m)$ are removed at the end of the algorithm. Also, (58) is a consequence of (25) and (59) is satisfied by choosing $\delta < \frac{\gamma}{16P+2\gamma}$. The last inequality is secured by the appropriate choice of n_0 . The above inequality shows the conflict of the power constraint and the assumption that (24) can hold for some $n > n_0$, where n_0 is chosen such that for any $n > n_0$

$$\frac{1}{\delta} e^{-(s(n)-nR_{FB})} < \frac{1}{2} \quad (61)$$

$$\frac{2}{\delta} e^{-n\epsilon} < \frac{1}{2}, \quad (62)$$

$$e^{-n(R-R_{FB})} < \frac{\gamma}{16P + \gamma}. \quad (63)$$

Given the assumption of $R_{FB} < R$, it is clear that there exists n_0 such that all the above three inequalities hold and this completes the proof. \blacksquare

Proof of Theorem 4: We prove the achievability of the above error exponent using an iterative scheme similar to the one used in the proof of Theorem 1. We use the exact same structure and notation as in the previous iterative scheme and just express the distinctions of this scheme. The main distinction is that here, instead of feeding back the decoded message (i.e. $U_{n_1} = \hat{m}_1$), the receiver sends back a function of its decoded message

$$U_{n_1} = g^{(n)}(\hat{m}_1), \quad (64)$$

where $g^{(n)} : \mathcal{M} \mapsto \{1, \dots, e^{nR_{FB}}\}$ is the feedback decision function. After receiving U_{n_1} , the transmitter compares the received feedback with the feedback corresponding to the original message and stays silent if

$$g^{(n)}(m) = U_{n_1}.$$

Otherwise, it sends the failure alarm and retransmits the message with high power exactly similar to what was described in the proof of Theorem 1. Considering the range of the feedback function $g^{(n)}(\cdot)$, it is

clear that this scheme meets the feedback constraint. Also it is easy to show that the power constraint is also met. In particular, note that the probability of retransmission in our scenario is

$$\Pr\{g^{(n)}(m) \neq g^{(n)}(\hat{m}_1)\}$$

which is less than or equal to $\gamma = \Pr\{m \neq \hat{m}_1\}$ and therefore the expected power used here is less than the case considered in Theorem 1. Also note that the types of errors seen here include the three types of errors in the earlier case (false negative, false positive and wrong decoding at the receiver) plus the error due to the fact that a subset of the decoding errors in the first block are not recognized by the transmitter. That is, the error corresponding to the event

$$\{m \neq \hat{m}_1, g^{(n)}(m) = g^{(n)}(\hat{m}_1)\},$$

which we call an *error mis-detection event*, must also be considered as a possible error event. We showed earlier that the algorithm in Theorem 1 achieves a doubly exponential error decay, where the error is associated with the first three types of errors. Therefore, the probability of error for the current scenario can be upper bounded by the sum of two terms: the probability associated with an error mis-detection event and the probability associated with the other three types of errors:

$$P_e(n, R, R_{FB}, P) \leq \Pr\{m \neq \hat{m}_1, g^{(n)}(m) = g^{(n)}(\hat{m}_1)\} + \exp(-\exp(n(\zeta + o(1)))), \quad (65)$$

for some $\zeta > 0$. Given that the feedback rate is less than the feedforward rate, we expect the error mis-detection event to dominate the total error probability. Hence, the proof will be complete if we show that there exists a sequence of feedback encoding functions $\{g^{(n)}(\cdot)\}_{n=1}^\infty$ such that

$$\Pr\{m \neq \hat{m}_1, g^{(n)}(m) = g^{(n)}(\hat{m}_1)\} \leq \exp(-n(E_{NoFB}(R) + R_{NoFB} + o(1))). \quad (66)$$

We show the existence of such a feedback encoder sequence using a random coding argument. Given n and a feedback function $g^{(n)} : \mathcal{M} \mapsto \{1, \dots, e^{nR_{FB}}\}$, let's define the set $\mathcal{V}^{(n)}(j)$ for each $j \in \{1, \dots, e^{nR_{FB}}\}$ as

$$\mathcal{V}^{(n)}(j) = \{m \in \mathcal{M} : g^{(n)}(m) = j\}.$$

We can observe that, in fact, determining the function $g^{(n)}(\cdot)$ is equivalent to partitioning $\{1, \dots, e^{nR}\}$ into the sets $\{\mathcal{V}^{(n)}(j)\}_{j=1}^{e^{nR_{FB}}}$. Now let's consider all the possible feedback functions for which

$$|\mathcal{V}^{(n)}(j)| = e^{n(R-R_{FB})},$$

for all $j \in \{1, \dots, e^{nR_{FB}}\}$. That is, let's consider all the equal partitionings of the set $\{1, \dots, e^{nR_{FB}}\}$. From this set of functions, let's pick the function $g^*(\cdot)$ uniformly randomly and use it as the feedback encoder function. We denote the partitioning associated with $g^*(\cdot)$ by $\{\mathcal{V}^*(j)\}_{j=1}^{e^{nR_{FB}}}$. Now let's compute

$$E[\Pr\{m \neq \hat{m}_1, g^{(n)}(m) = g^{(n)}(\hat{m}_1)\}],$$

where the expectation is with respect to the randomness in picking the feedback function. We have

$$\begin{aligned} & E[\Pr\{m \neq \hat{m}_1, g^*(m) = g^*(\hat{m}_1)\}] = \\ & E[\sum_{m=1}^{e^{nR}} \Pr\{m \text{ is sent}\} \sum_{i \in \mathcal{M}, i \neq m} \Pr\{\hat{m}_1 = i | m \text{ is sent}\} \mathbf{1}_{\{g^*(i) = g^*(m)\}}] = \\ & \sum_{m=1}^{e^{nR}} \Pr\{m \text{ is sent}\} \sum_{i \in \mathcal{M}, i \neq m} \Pr\{\hat{m}_1 = i | m \text{ is sent}\} E[\mathbf{1}_{\{g^*(i) = g^*(m)\}}] . \end{aligned} \quad (67)$$

For each pair (i, m) , we can write

$$\begin{aligned} E[\mathbf{1}_{\{g^*(i) = g^*(m)\}}] &= \Pr\{g^*(i) = g^*(m)\} \\ &= \sum_{k=1}^{e^{nR_{FB}}} \Pr\{g^*(i) = k | g^*(m) = k\} \Pr\{g^*(m) = k\} \\ &= \sum_{k=1}^{e^{nR_{FB}}} \Pr\{i \in \mathcal{V}^*(k) | m \in \mathcal{V}^*(k)\} \Pr\{m \in \mathcal{V}^*(k)\}. \end{aligned} \quad (68)$$

Since $\{\mathcal{V}^*(j)\}_{j=1}^{e^{nR_{FB}}}$ is uniformly randomly chosen from all equal partitionings of $\{1, \dots, e^{nR}\}$, we can write for $i \neq m$ and for any $k \in \{1, \dots, e^{nR_{FB}}\}$

$$\begin{aligned} \Pr\{i \in \mathcal{V}^*(k) | m \in \mathcal{V}^*(k)\} &= \frac{|\mathcal{V}^*(k)| - 1}{\sum_{k'=1}^{e^{nR_{FB}}} |\mathcal{V}^*(k')| - 1} \\ &= \frac{e^{n(R-R_{FB})} - 1}{e^{nR} - 1} \end{aligned}$$

Substituting the above equality in (68) we get

$$E[\mathbf{1}_{\{g^*(i)=g^*(m)\}}] = \frac{e^{n(R-R_{FB})} - 1}{e^{nR} - 1}. \quad (69)$$

We can now combine (69) and (67) and conclude

$$\begin{aligned} E[\Pr\{m \neq \hat{m}_1, g^*(m) = g^*(\hat{m}_1)\}] &= \frac{e^{n(R-R_{FB})} - 1}{e^{nR} - 1} \sum_{m=1}^{e^{nR}} \Pr\{m \text{ is sent}\} \sum_{i \in \mathcal{M}, i \neq m} \Pr\{\hat{m}_1 = i | m \text{ is sent}\} \\ &= e^{-n(R_{FB}+o(1))} \Pr\{\text{Decoding error in first block}\} \\ &\leq e^{-n(R_{FB}+o(1))} P_e(n, R, P) \\ &\leq e^{-n(R_{FB}+E_{NoFB}(R)+o(1))}. \end{aligned}$$

The above inequality implies that the expected (with respect to encoder selection) probability of error mis-detection event is less than the right hand side of (66). Therefore, we can conclude that there exists at least one feedback encoding function among the ones from which we randomly selected that satisfies (66). This completes the proof. \blacksquare

Proof of Theorem 5: Here, we only present the proof for the case where $L = 2$. Following a similar approach as in Theorem 2, the proof can be extended to $L > 2$.

For each $R < C$, there exists $\delta' > 0$ such that $R < C(1 - \delta')$. Let's fix δ' and consider the integer k which satisfies

$$\frac{\delta'}{2} \leq \frac{1}{k} < \delta'. \quad (70)$$

We divide the whole transmission block into k sub-blocks each with length $l = n/k$. We then partition each sub-block into three parts of lengths l_1 , 1 and l_2 exactly the same as the partitioning in the 3-phase scheme proposed in Section III. In the first portion of sub-block $j \in \{1, \dots, k-1\}$, message m_j which contains $nR/(k-1)$ nats of new information is transmitted on the forward channel using a non-feedback Gaussian codebook similar to the first phase of the algorithm described in Section III. After the transmission, this message is decoded and the decoded message \hat{m}_j is transmitted back on the feedback channel during the first portion of the $j+1^{st}$ sub-block and with the rate R nats per channel use. By the end of the feedback transmission (end of the first portion of sub-block $j+1$), the transmitter can detect the decoding error. If $\hat{m}_j \neq m_j$, the failure alarm is sent in the second portion of the $j+1^{st}$ sub-block and the message m_j is retransmitted with high power in the third portion of the $j+1^{st}$ block. In fact, for each sub-block we apply the 3-phase iterative scheme of Section III with the distinction that the error detection and retransmission for each sub-block occurs one sub-block after the original transmission. The forward rate per channel use in each sub-block is

$$\frac{kR}{k-1} < C(1 - \delta')(1 - \frac{1}{k}) < C(1 - \delta')^2.$$

Defining $\delta = 2\delta'$, the rate per channel use will be less than $C(1 - \delta)$. Using the results of Section III, we can conclude that there exists $\zeta > 0$ such that the error probability P_e^j for the message m_j is upper bounded by

$$P_e^j < \exp(-\exp(\frac{n}{k}\zeta)) \leq \exp(-\exp(n\frac{\delta'\zeta}{2})),$$

where the last inequality is a consequence of (70). Using the union bound, the total error probability will be bounded as follows

$$\begin{aligned} P_e &\leq \sum_{j=1}^{k-1} P_e^j \\ &\leq (k-1)\exp(-\exp(n\frac{\delta'\zeta}{2})) \\ &\leq \frac{2}{\delta'}\exp(-\exp(n\frac{\delta'\zeta}{2})), \end{aligned}$$

where the last inequality is again a consequence of (70). Taking $\theta = \frac{\delta'\zeta}{2}$, the above inequality completes the proof. ■

REFERENCES

- [1] C.E.Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Systems Technical Journal*, vol. 38, no. 3, pp. 611-656, 1959.
- [2] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," *IEEE Trans. Info. Theory*, vol. IT-12, pp. 172-182, 1966.
- [3] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," arXiv:0812.2709v2, 16 August 2009.
- [4] Y.-H. Kim, A. Lapidoth, and T. Weissman, "The Gaussian channel with noisy feedback" in *Proc. of the International Symposium on Information Theory (ISIT-07)*, (Nice, France), pp. 1416-1420, June 2007.
- [5] Young-Han Kim, Amos Lapidoth and Tsachy Weissman, "Error Exponents for the Gaussian Channel with Active Noisy Feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1223-1236, March 2011.
- [6] James M. Ooi, "A framework for low-complexity communication over channels with feedback," Ph.D. dissertation, MIT, 1997.
- [7] A. J. Kramer, "Improving communication reliability by use of an intermittent feedback link," *IEEE Transactions on Information Theory*, vol. IT-15, pp. 52-60, January 1969.
- [8] Nuno C. Martins and Tsachy Weissman, "Coding for Additive White Noise Channels With Feedback Corrupted by Quantization or Bounded Noise," *IEEE Transactions on Information Theory*, Volume 54, Issue 9, Sept. 2008 Page(s): 4274-4282
- [9] M. Agarwal, D. Guo and M. Honig, "Error Exponent for Gaussian Channels with Partial Sequential Feedback", *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.
- [10] S. C. Draper, K. Ramchandran, B. Rimoldi, A. Sahai, and D. N. C. Tse, "Attaining maximal reliability with minimal feedback via joint channel-code and hash-function design," in *Proc. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2005.
- [11] S. C. Draper and A. Sahai, "Noisy feedback improves communication reliability," in *Proc. IEEE International Symposium on Information Theory*, Seattle, WA, July 2006, pp. 69-73.
- [12] K. Eswaran, A. Sarwate, A. Sahai, and M. Gastpar, "Zero-rate feedback can achieve the empirical capacity," November 2007, submitted to *IEEE Transactions of Information Theory*.
- [13] O. Shayevitz and M. Feder, "Optimal Feedback Communication via Posterior Matching," *IEEE Transactions on Information Theory*, Vol. 57, No. 3, pp. 1186-1222, March 2011.
- [14] T. P. Coleman, "A Stochastic Control Viewpoint on 'Posterior Matching-Style Communication Schemes'," *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.
- [15] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Probl. Inf. Transm.*, vol. 4, no. 4, pp. 319, 1968.